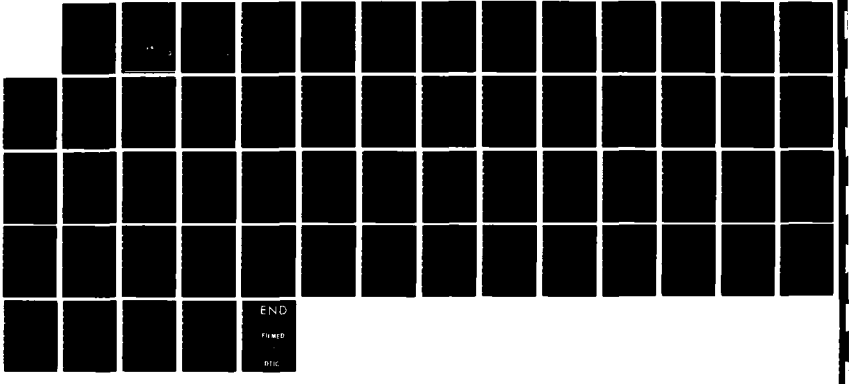
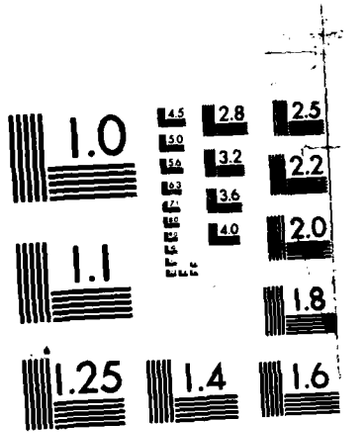
AD-A161 754 THE RELATIONSHIP BETWEEN COGNITIVE CATEGORIES OF RATERS AND RATING ACCURACY(U) MICHIGAN STATE UNIV EAST LANSING C OSTROFF ET AL OCT 85 TR-85-4 N00014-83-K-0756 1/1

UNCLASSIFIED F/G 5/10 NL





(12)

MICHIGAN STATE UNIVERSITY

Industrial/Organizational Psychology
and
Organizational Behavior

The Relationship Between Cognitive Categories of Raters and Rating Accuracy

Cheri Ostroff and Daniel R. Ilgen

Michigan State University

Michigan State University
East Lansing, Michigan 48824

The Relationship Between Cognitive Categories of Raters and Rating Accuracy

Cheri Ostroff and Daniel R. Ilgen

Michigan State University

Prepared for
Office of Naval Research
Organizational Effectiveness Research Programs
Code 4420E

Grant No. N00014-83-K-0756
NR170-961

Technical Report No. 85-4
Department of Psychology
and
Department of Management
Michigan State University

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: 85-4	2. GOVT ACCESSION NO.: AD-A161 754
	3. RECIPIENT'S CATALOG NUMBER:
4. TITLE (and Subtitle): The Relationship Between Cognitive Categories of Raters and Rating Accuracy	5. TYPE OF REPORT & PERIOD COVERED: Interim
	6. PERFORMING ORG. REPORT NUMBER: 2009
7. AUTHOR(s): Cheri Ostroff and Daniel R. Ilgen	8. CONTRACT OR GRANT NUMBER(s): N00014-83-K-0756
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Psychology, Michigan State University, East Lansing, MI 48824-1117	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NR170-961
11. CONTROLLING OFFICE NAME AND ADDRESS: Organizational Effectiveness Research Programs, Office of Naval Research (Code 4420E), Arlington, VA 22217	12. REPORT DATE: October, 1985
	13. NUMBER OF PAGES: 51
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Performance ratings, personal constructs, cognitive categories, halo, cognitive differentiation

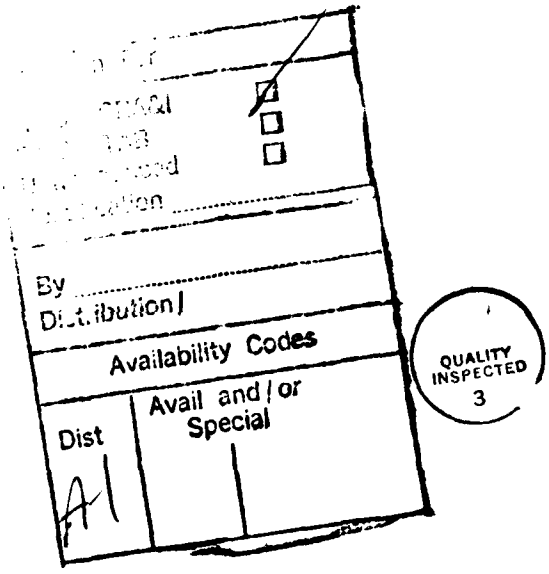
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This research investigated cognitive processes affecting the accuracy of performance evaluation. One hundred and twenty-five hospital nurses completed questionnaires measuring their cognitive categorization processes and then rated a videotaped nurse's performance. Results focusing on the match between raters' cognitive categories and rating scales indicated that: a) rating accuracy was related to the match between raters' cognitive category dimensions and rating scale dimensions, but b) not related to the match between rating scales and raters general tendency to describe ratees in terms of behaviors or personal characteristics

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(traits). Little support was found for the effect of misclassifying behaviors and dimensions irrelevant to job performance on accuracy. Raters' ability to differentiate clearly between rating dimensions was related to accuracy and to halo in ratings. Finally, work experience and job position were related to raters' cognitive systems while prior rating experience was important for accuracy. Limitations of the study and implications of future research investigating the cognitive processes of raters in relation to rating accuracy are discussed.



The Relationship Between Cognitive Categories of Raters and Rating Accuracy

According to the cognitive processing view, appraising performance involves gathering, storing and recalling information (Cooper, 1981; Feldman, 1981; Ilgen & Feldman, 1983; Landy & Farr, 1980). Central to this view is the categorization of information into dimensional schemata (e.g. Ilgen & Feldman, 1983). Theoretical explications of the role of cognitive categories in processing performance information and their effects on accuracy and errors are numerous, but little empirical work exists (Nathan & Lord, 1983). The present study examines raters' category sytems in relation to the accuracy of their performance evaluations.

Two bodies of research are relevant to the effects of *categories on appraisal accuracy:* implicit personality theory and personal construct theory. Each of these will be briefly addressed before presenting specific hypotheses.

Implicit Personality Theory

Implicit personality theory is concerned with how individuals' believe traits covary (Bruner & Tagiuri, 1954; Schneider, 1973). It has been shown that raters use their own categories, or implicit theories, to judge others and that these categories relate to trait dimensions (e.g. Passini & Norman, 1966). Thus, the rater's beliefs about trait covariations affect the evaluation of others (Hakel, 1969; Norman & Goldberg, 1966; Passini & Norman, 1966).

Since raters may possess implicit "theories" about trait dimensions and intercorrelations among these dimensions which may or may not match actual conditions, raters' implicit theories may have important implications for rater accuracy (Nathan & Alexander, 1985). Raters whose implicit theories about performance closely match the ratee's actual performance are more likely to provide accurate ratings than those whose implicit assumptions about behavior are inconsistent with actual performance (Borman, 1983; Landy & Farr, 1980; Nathan & Alexander, 1985).

Implicit personality theory has been used to explain two rating errors, halo and systematic distortion. Halo errors result in artifactually high intercorrelations among performance dimensions. When comparing intercorrelations of ratings with known covariances among performance dimensions, halo errors were found suggesting that individuals distort the magnitude of relationships between dimensions of personality and job performance (Borman, 1975; Nisbett & Wilson, 1977). Systematic distortion reflects the tendency to overestimate the degree of correlation between dimensions that are semantically similar, such as interpersonal skills and verbal fluency. Shweder and D'Andrade (1980) found that either the absence of relevant information about ratees or time delays between observations and rating led to inter-dimension correlations of ratings which were biased in the direction of semantic similarity.

Most work on implicit personality theory related to performance ratings simply demonstrates that errors are consistent with implicit theories. What is needed at this point is an assessment of the theories people use. Personal construct theory provides a basis for addressing the theories used by people by exploring individual differences in cognitive category systems relevant to person perception.

Personal Construct Theory

In his personal construct theory, Kelly (1955) asserted that each individual formulates, in his own way, constructs through which he or she views the world of events. That is, individuals develop personal construct systems, or categories, which they use to judge people and events. While similar to implicit personality theory in that both theories postulate interpersonal "filtering" of information by perceivers, personal construct theory examines individual differences in these filters in terms of their structure and content, while implicit personality theory focuses on the covariance of traits in raters' category systems (Borman, 1983).

Most research in personal construct theory has used the Role Construct Reporatory Test (RCRT). This test requires respondents to record names of persons who fit a number of roles. The respondent is then asked to consider various triads of these role persons, and for each triad, identify an important way in which two of the persons are alike, yet different from the third. Taken together, the responses constitute measures of the person's

personal constructs. Studies utilizing the RCRT have shown that individuals prefer to use their own constructs to rate others (Bonarius, 1965), they differentiate more finely between ratees when employing their own constructs (Adams-Webber, 1979; Isaacson, 1966) and the content of individuals' constructs differs across people (Sechrest, 1968; Rosenberg, 1977). Yet, none of this work focused upon performance appraisals. Research is needed to assess the impact of individual differences in categories on observations of work behavior and on performance ratings.

Role of Categorization in Performance Ratings

Some research has addressed more directly the effect of raters' categories on performance ratings. Nathan & Lord (1983) compared Borman's (1978) notion that raters store information in independent dimensions with that of Feldman's (1981) which assumes that information is automatically stored and integrated. Results indicated that Borman's model was useful in demonstrating raters' ability to differentiate between performance dimensions; however, the presence of a large halo effect was consistent with Feldman's model. The authors concluded that the data supported both models, perhaps due to individual differences in cognitive styles of raters.

Cognitive complexity has been suggested as an individual difference characteristic relevant to information processing related to performance (Feldman, 1981; Kane & Lawler, 1979; Landy & Farr, 1980). Cognitive complexity is the "degree to which a person

possesses the ability to perceive behavior in a multidimensional manner" (Schneier, 1977; p. 541). Bernardin, Cardy & Carlyle (1982) proposed that in an appraisal situation, cognitive complexity should be reflected in the persons' ability to conceptualize performance into multiple dimensions. However, results from studies investigating the relationship between cognitive complexity and rating errors, acceptance of the format, confidence in ratings, or accuracy are mixed (Bernardin, Cardy & Carlyle, 1982; Borman, 1979; Lahey & Saal, 1981; Sauser & Pond, 1981).

Finally, some research has focused upon the actual content of cognitive categories. Since performance appraisal instruments typically stress using behavior rather than trait dimensions, it is important to know whether people tend to encode observations into behavior rather than trait dimensions. Evidence suggests that this behavioral information is integrated into cognitive categories which are global and/or trait-based, rather than based on the specific behaviors observed (Murphy, Martin & Garcia, 1982). Thus, while performance rating instruments typically require raters to focus on job behaviors, the effect of observing these behaviors and then incorporating them into the category systems of raters may seriously bias the ratings.

Taken together, the research indicates that individual differences in raters' category systems do exist and that the categories themselves influence performance ratings. At this

point, it is not clear exactly what these category systems are, how they are structured, how consistent they are with the rating scales provided in performance appraisal, or how they impact performance evaluations, other than to suggest that all of these issues affect the degree to which accurate ratings of employees can be made.

Objectives of the Present Study

This research explores the nature of category systems and the affects of these systems on the accuracy of performance ratings. Also addressed are possible predictors or correlates of specific category systems. The following hypotheses were tested.

Category to Scale Match. Raters should provide more accurate ratings to the extent that their personal category systems match those defined on the rating forms used in appraising performance. Thus, it follows that:

Hypothesis One: When using behaviorally-based rating scales, raters possessing behaviorally-oriented category systems will yield more accurate ratings than raters with trait-like categories. For trait-based scales, raters possessing trait-oriented category systems will yield more accurate ratings than raters with behaviorally-based categories.

Although as suggested in Hypothesis One, ratings should be more accurate if the general nature of the category system matches the general nature of the scale, a more refined level of match involves the specific dimensions and behaviors relevant to the job being rated. Raters trained to recognize the specific job

dimensions and their corresponding behaviors have consistently outperformed those who have not received such training (Bernardin & Pence, 1980; McIntyre, Smith & Hassett, 1984; Pulakos, 1984, in press). Underlying this training is the assumption that accuracy is increased because raters have developed a category system that matches the performance rating scale. A similar notion is reflected in Hypothesis Two:

> Hypothesis Two: To the extent that raters are able to dimensionalize job behaviors in a manner consistent with that of the rating scale, ratings will be more accurate.

Behavioral Differentiation. In appraising performance, the rater must determine which of the ratee's behaviors are job-related and which are not. Yet, considerable evidence suggests that non-performance related characteristics and behaviors of the ratee (i.e. sex, race, etc.) are observed and serve to bias ratings (Ilgen & Feldman, 1983; Landy & Farr, 1980). This implies that:

> Hypothesis Three: Accuracy in ratings will be related to the degree to which a rater is able to distinguish between behaviors and dimensions that are relevant to job performance and behaviors and dimensions that are irrelevant to job performance.

Cognitive Differentiation. When considering the work situation, if raters are able to differentiate behaviors into dimensions with little degree of overlap, their ratings should be more accurate. Specifically:

Hypothesis Four: More accurate raters have highly differentiated category systems for the job such that low intercorrelations exist between category dimensions, while less accurate raters are unable to differentiate clearly between dimensions.

Similarly, the cognitive differentiation of raters should be related to the degree of halo in ratings (Schneier, 1977). Hypothesis Five reflects this notion.

Hypothesis Five: Raters with more highly differentiated category systems for the job will exhibit less halo in their ratings than those with less differentiated systems.

Experience. Category systems are learned (Rosch, Mervis, Gray, Johnson, & Boyers-Braem, 1976). Furthermore, if we assume that those who are promoted learn more about the organization from these experiences, such experiences should influence their cognitive categories. Hypothesis Six is based on this rationale.

Hypothesis Six: Rater experience will be correlated with the category system he or she uses to evaluate others and with rating accuracy.

Method

Overview

The research was conducted in two phases. The first phase involved the development of instruments needed for measuring relevant variables and the filming of a videotape with the properties necessary for the rating stimulus. A number of

different samples participated in this phase. In the second phase, nurses viewed the videotape and completed the research measures at the hospitals where they were employed.[1]

Development of Questionnaire Measures and Stimulus Materials

Questionnaire Measures. Three sets of measures were developed. These were: the Role Grid and the Behavior Grid which were designed to assess category systems; a Background Questionnaire which assessed possible correlates of category systems; and two rating scales with corresponding true score ratings to assess rating accuracy (For a more thorough description of each measure, see Ostroff, 1985.)

Role Grid. The Role Grid, based on Kelly's reporatory grid technique, assessed the degree to which nurses possessed trait-based or behaviorally-based category systems. The grid presented triads of job roles. People were asked to: 1) select two job roles in a triad which they felt were similar, and 2) describe, in writing, how they believed the two roles were similar.

To develop the triads, pairs of job roles were presented to a sample of five nurses and eight graduate students who described how the two roles were similar. From a large list, the first criterion for retaining roles in a triad was to have at least 70% of the sample identify a trait for two roles and a behavior for another pair in the triad. A second sample of 15 nurses responded only to those triads that met the 70% criterion, and triads were eliminated if at least 33% of the people were unable to identify either a

behavior or trait construct for the triad. Eight triads were retained in the final version of the grid, each of which contained two roles frequently seen as sharing behaviors and two sharing traits. For example, consider a triad of artist, comedian and cartoonist. The artist and comedian could be seen as sharing the behavior of drawing while the comedian and cartoonist might possess the trait of humor. Figure 1 is a sample of Role Grid.

Behavior Grid. The Behavior Grid was developed to assess the extent to which raters were able to correctly identify ratee behaviors which belonged to particular dimensions of the job, the extent to which behaviors irrelevant to the job were likley to be seen as relevant, the extent to which job relevant behaviors were viewed as irrelevant to job performance dimensions, and the extent to which raters differentiated between behaviors and dimensions. The form of the final scale is displayed in Figure 2. Note that the rows of the grid are behaviors of two types--behaviors believed to be relevant to performance of a nurse (i.e., "this nurse could not be expected to observe that a patient consistently leaves untouched a particular type of food") and irrelevant to job performance (i.e., "would expect to find this nurse exercising, jogging, or working out during her/his breaks or free time"). The columns represented dimensions and were also of two types--job related dimensions (i.e., Observational Ability) and non-job related dimensions (i.e., Sense of Humor). Placement of items within rows and columns was random. Nurses were instructed to

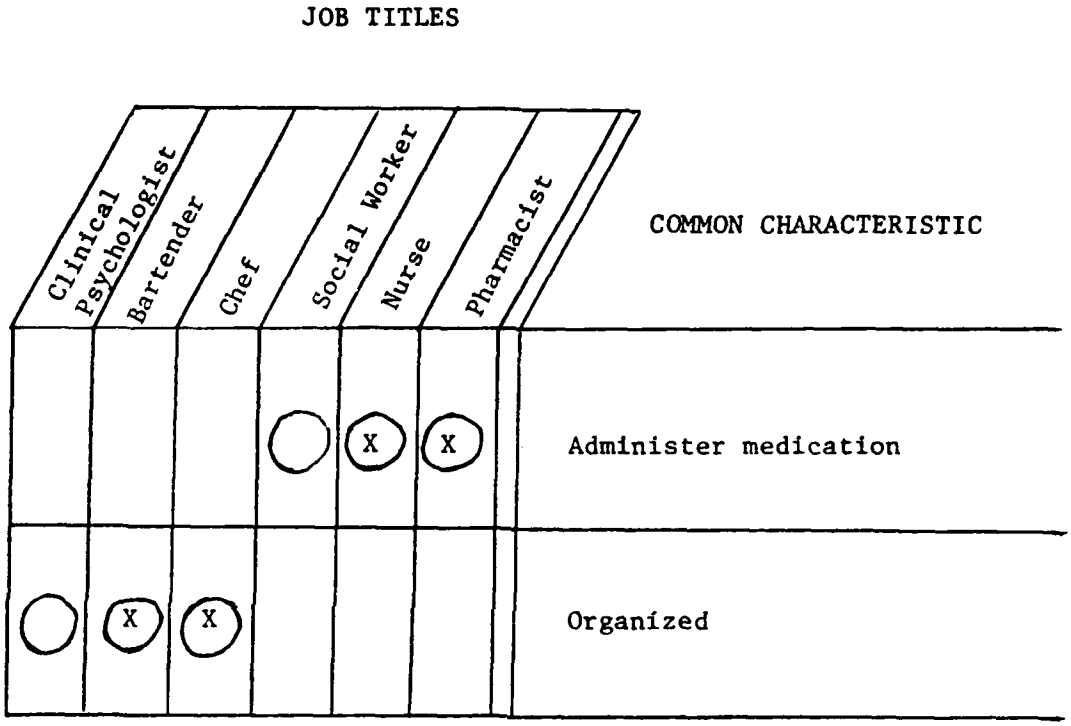


Figure 1. Sample of Role Grid.

DIMENSIONS

Organizational Ability	Knowledge and Judgment	Appearance	Family-Oriented	Skill in Human Relations	BEHAVIORS
				X	This nurse could be expected, whenever possible, to sit down and talk with a terminal-cancer patient who is considered to be "demanding".
			X		Would expect this nurse to change her/his hairstyle every few months
	X				Would expect this nurse to know enough to delay giving regular insulin to a patient who was to have a fasting blood sugar, until after the blood had been drawn.

Figure 2. Sample of Behavior Grid.

consider each behavior (row) and place a check under the column(s) where they felt the behavior belonged.

Items on the Behavior Grid were selected using the translation-retranslation method of Smith and Kendall (1963) in the development of Behaviorally Anchored Rating Scales (BARS). The initial sample of job relevant behaviors and dimensions were those used on the original Smith and Kendall scale developed for nurses. Non-job performance behaviors and dimensions were generated from critical incidents supplied by a sample of five nurses. Sample items unrelated to performance are: dresses fashionably, smiles a lot, and calls spouse while at work. Seven nurses retranslated the pool of job relevant and 15 graduate students the non-job relevant behaviors. A final set of 20 job related and 20 non-job related behaviors were sorted into the dimensions with at least 87% agreement among the raters resulted.

Variables Measured on Role and Behavior Grids

The written response to each item on the Role Grid was coded as either "Behavior," "Trait" or "Other" by the experimenter. To ensure objectivity and reliability of the coding of the written responses, an independent scorer coded two separate samples of the Role Grid. The experimenter and the independent scorer agreed on 90% of the codings for the first sample and 89% for the second. Due to the high level of agreement, only the experimenter's codings were used. Once coded, the following measures were derived from the Role Grid:

1. Behavior. The number of pairs which were seen as sharing a common behavior.
2. Trait. The number of pairs seen as sharing a common trait.
3. Other. The number of times neither a behavior nor a trait was viewed for a pair.

For a sample of 8 head nurses and 11 undergraduate students who were administered the Role Grid twice, with approximately a one month delay, the test-retest reliabilities were: .83 for Behavior; .85 for Trait; and .70 for Other. Although these are quite acceptable reliabilities, keep in mind they are not independent due to the ipsative nature of the scale.

For the Behavior Grid, six variables were constructed. These were:

1. Rating Scale Similarity. From the subset of dimensions on the Behavior Grid which were identified a priori as relevant to the nurse's job and a subset of job relevant behaviors that described those dimensions, each behavior was scored on a scale ranging from 6 to 1 depending on the degree to which the response matched the BARS scale. For example, a score of six (perfect match) occurred if the behavior was correctly placed in the appropriate job dimension; a score of 4 indicated placement in the correct dimension but also placement in two other job dimensions; a score of 1 indicated incorrect placement.

The index was the sum of the scores for these job related behaviors and ranged from 20 to 120. High scores indicated a greater match to the BARS scale.

2. Non-Job Relevant Behavior Classification. This index was the sum of the number of times non-job relevant behaviors were misclassified as belonging to job relevant dimensions.

3. Job Relevant Behavior Classification. In a manner similar to 2 above, the number of times behaviors identified as job relevant were misclassified as belonging to non-job relevant dimensions was tallied.

4. Overall Cognitive Differentiation. This index was computed by totalling the number of check marks (or number of times behaviors were placed in dimensions) each rater placed in the grid. Low scores indicated a greater tendency to differentiate behaviors into dimensions.

5. Job Behavior Cognitive Differentiation. This index was computed in a manner similar to 4 above, but only for the job related behaviors in the grid.

6. Non-Job Behavior Cognitive Differentiation. In a manner similar to 4 above, the number of check marks each rater placed in the grid for non-job related behaviors was tallied.

Eleven head nurses completed the Behavior Grid on two occasions, one month apart. For each nurse, the percent of

responses in the grid that remained the same over the two administrations of the scale was determined. These percentages ranged from 85% to 97%; the average percentage of unchanged responses over time was 92%.

Background Questionnaire. A Background Questionnaire was developed to measure basic demographic and background variables of the nurses which may affect nurses' schemas and rating accuracy. The items in this questionnaire included years of experience on the job, job position, job title, unit in the hospital, educational experience, highest educational degree, sex and experience with rating. Table 1 presents the percentage of nurses falling in each level of each experience variable.

Performance Rating Scales. Five of the original dimensions from the BARS scale developed by Smith and Kendall (1963) were used by nurses to rate the videotaped performance of a nurse. The five dimensions were Knowledge and Judgment, Organizational Ability, Skill in Human Relations, Conscientiousness, and Observational Ability.[2] A trait-based rating scale was also developed for use when rating the videotaped nurse's performance. The traits used in the scale were culled from previously developed rating scales for nurses. The scale contained six trait dimensions with a short definitional description of each and a five point Likert-type scale ranging from exceptional to unsatisfactory. The six trait dimensions were Compassionate, Helpful, Proficient, Efficient, Communicative and Perceptive.

Table 1

Percent of Nurses in Each Level of the Experience Variables

Variable	Percent
Years worked as nurse:	
less than 1 year	1%
1 to 4 years	8%
5 to 10 years	30%
11 to 20 years	35%
21 to 30 years	18%
over 30 years	9%
Position:	
Staff Nurse	3%
Charge Nurse	35%
Head Nurse	22%
Supervisor	16%
Other	24%
Title:	
Licensed Practical Nurse	1%
Registered Nurse	91%
Nurse Practitioner	1%
Other	6%
Sex:	
Male	3%
Female	97%

(table continued)

Table 1 (continued)

Percent of Nurses in Each Level of the Experience Variables

Variable	Percent
Unit working in Hospital:	
Intensive Care	11%
Emergency	4%
Geriatrics	0%
Surgery	12%
Psychiatric	3%
OB/GYN	17%
Medical	13%
Children	2%
Other	37%
Educational Training:	
Community College (2 years)	33%
Hospital (3 years)	47%
College (4 years)	20%
Highest Educational Degree:	
Associate Degree	60%
Bachelor's Degree	32%
Master's Degree	8%
Ph.D.	0%

(table continued)

Table 1 (continued)

Percent of Nurses in Each Level of the Experience Variables

Variable	Percent
Prior Rating Experience:	
Yes	87%
No	13%
Years of Rating Experience:	
No experience	13%
less than 1 year	5%
1 to 4 years	33%
5 to 10 years	25%
over 10 years	23%

Videotape

A 25 minute videotape featuring a nurse in a hospital setting served as the stimulus material for ratings. The tape featured 18, one to three minute, scenes depicting enactments of job behaviors from one or more of the five performance dimensions.

To develop the scenes, behavioral examples for each job dimension from the BARS scale were modified by the experimenter and two nurses. Within each dimension, the ratee's behavior was designed to be consistent in performance level, but across job dimensions, the performance level was varied. For three dimensions, the ratee exhibited examples of good performance; on one dimension, the ratee exhibited average performance; and one, poor performance. The scenes were randomly ordered in the final videotape. Trait dimensions were also exhibited on the videotape. The behavior and trait dimensions represented appear in Table 2.

Two sets of expert raters (10 graduate nursing students for the BARS and 10 for the Trait scale) viewed and evaluated each scene on the videotape for two purposes. First, their ratings were used to eliminate scenes that did not produce agreement among raters as to: a) the performance dimensions and/or trait dimensions represented in the scenes, or b) the effectiveness level of the behavioral/trait dimension represented. For the 18 scenes retained, interrater reliabilities for the assignment of scenes to dimensions ranged from .73 to .95 for the BARS dimensions and from .78 to .91 for the trait dimensions. Cronbach's generalizability

Table 2

True Score Ratings of Performance for the BARS and Trait Rating Scales

Performance Dimension	True Score Mean	SD
BARS Scale		
Knowledge and Judgment	1.38	.189
Organizational Ability	0.19	.088
Skill in Human Relations	1.66	.223
Conscientiousness	1.59	.249
Observational Ability	1.56	.125
Trait Scale		
Compassionate	4.18	.459
Helpful	4.01	.369
Proficient	3.78	.382
Perceptive	4.01	.283
Communicative	3.54	.693
Efficient	1.05	.071

Note. Means and SD's are based on a 9-point rating scale ranging from 0.0 to 2.0 in units of 0.25 for BARS scale, and a 5-point rating scale ranging from 1.0 to 5.0 in units of 1.0 for trait scale.

coefficient, using scenes and dimensions as fixed factors and raters as random, was .94 for behavior dimensions and .96 for trait dimensions.

Once a set of scenes was identified which met the inclusion criteria, the expert raters' ratings of the individual scenes were used as the standard or true scores to which the nurse subjects' ratings were compared and from which the performance accuracy indices were computed. True scores for each dimension were derived by averaging the mean rating scores for the scenes which were identified as representing the dimension. The true score means for the BARS and trait scale dimensions appear in Table 2.

Criterion Measures

Accuracy. Four accuracy measures, two for the BARS scale and two for the Trait scale, were calculated. For each rater, Cronbach's (1955) component of overall accuracy was computed by squaring the difference between the rated and true scores and summing over all dimensions. Lower overall accuracy scores indicated greater accuracy.

For each rater, correlational accuracy was computed by correlating the true scores and the observed scores, for the BARS and also for the Trait scale. Higher correlational accuracy scores indicated greater accuracy in terms of the pattern of performance levels across dimensions for the ratee.

Halo. Two measures of halo were computed, one for the BARS scale and one for the Trait scale. Halo was assessed as the standard deviation, within rater, of the ratings across dimensions.

Primary Study

Sample. Raters were 129 registered nurses, 125 females and 4 males, from three large midwestern hospitals. Ninety-two percent of the participants had five or more years of work experience and 87% had previous experience rating nurses' performance. Most (97%) were in some type of supervisory position. Four of the original sample were dropped due to missing data.

Procedure. For the primary study, nurses participated in a one and one-half hour long session and were assessed in groups of three to fifty persons per session. After a brief description of the project, nurses first completed the Background Questionnaire and then the Role Grid. Next they completed the Behavior Grid.

When all those in the session had completed the above measures, the questionnaires were collected. This was followed by explanation of the performance ratings scales, the videotape and the rating procedure. Nurses then viewed the videotape and rated the person on the tape using the BARS scale and the Trait scale. After these ratings, the nurses were debriefed and dismissed.

Results

Accuracy Measures

The means, standard deviations and intercorrelations for the nurses' accuracy scores--BARS overall accuracy, BARS correlational

accuracy, Trait overall accuracy and Trait correlational accuracy--are presented in Table 3. Within scale formats the accuracy scores were highly correlated (r = -.70 and r = -.61, for BARS and Trait scales respectively), but not between formats (r's ranged from -.18 to .47).

To test the relative accuracies of the BARS versus the Trait scale, it was first necessary to standardize the scale scores. The observed scores and the true scores for each dimension on each scale were transformed to z-scores before computing overall accuracy. Mean comparisons using t-tests revealed that for overall accuracy, nurses were significantly more accurate with the BARS scale than with the Trait scale ($t(1,125) = 2.82$, $p = .006$). In addition, the correlational accuracy score for each scale was transformed using Fischer's r-to-z transformation and a t-test was computed between the two means. No significant mean differences in accuracy were found for correlational accuracy scores ($t(1,123) = .60$, $p = .55$). Thus, it appears that participants were more accurate in discerning performance levels across dimensions when using the BARS scale than when using the Trait scale, but no difference in accuracy existed in nurses' ability to reflect the pattern of performance levels across dimensions (as reflected in correlational accuracy indices) when using the BARS or Trait scales.

One-way analyses of variance were performed for each of the accuracy measures by hospital groups to ensure the data could be

Table 3

Intercorrelations of Accuracy and Cognitive Measures

Variable	$\bar{X}$	SD	1	2	3	4	5	6	7	8	9	10	11	12
Accuracy														
1. BARS overall	8.74	8.57												
2. BARS correlational	0.94	0.25	-.70[a]											
3. Trait overall	10.92	6.16	.47	-.18										
4. Trait correlational	0.95	0.24	-.14	.20	-.61									
Category Orientation														
5. Behavior	4.40	2.19	-.02	.07	.07	.06								
6. Trait	3.04	2.12	.05	-.06	-.05	-.10	-.93							
Category Dimensionalizing														
7. Rating scale match	89.08	16.35	-.22	.27	-.12	.09	.05	-.09						
8. Non-job behavior classification	5.43	4.90	.10	-.10	.04	.04	-.14	.19	-.60					
9. Job behavior classification	2.48	3.02	.10	-.06	.06	.04	-.10	.13	-.57	.73				
Cognitive Differentiation														
10. Overall	76.67	36.21	.09	-.13	.02	.07	.11	.13	-.69	.85	.80			
11. Job	44.26	21.35	.07	-.08	.01	.09	-.06	.10	-.72	.80	.78	.97		
12. Non-job	32.51	16.19	.10	-.18	.04	.03	-.16	.18	-.60	.84	.75	.95	.86	

[a]Correlations above .15 are significant at $p \leq .05$ for one-tailed tests and .18 for two-tailed tests.

collapsed across the hospitals from which it was drawn. For three of the four measures, no differences existed across hospitals. For Trait overall accuracy, hospital means did differ ($F(2,121) = 3.10$, $p = .05$). Closer examination of these data, using Newman-Kuel's tests, revealed that the difference was due to one hospital in which accuracy scores on Trait overall accuracy were significantly lower than the other two.

Cognitive Measures

The means, standard deviations and intercorrelations for the cognitive processing indices also appear in Table 3. For the most part, the cognitive measures were highly intercorrelated. However, recall that the two measures from the Role Grid which assessed raters' category orientation, Behavior and Trait, were not independent; thus, their high intercorrelation was expected.

The Behavior and Trait measures revealed fairly low intercorrelations with the remaining cognitive measures (r's ranged from .05 to .19). The six cognitive measures derived from the Behavior Grid--Rating Scale Match, Job Behavior and Non-Job Behavior Classifications, and the three cognitive differentiation measures--were all highly intercorrelated (r's ranged from .57 to .97). These results suggest that the two grids may be measuring separate constructs. The Role Grid may measure category orientation of the raters while the Behavior Grid may assess the categorizing of behaviors and dimensions in the raters' cognitive

category system. It is also likely that common method variance contributed to the high intercorrelations within the Behavior Grid.

Analyses of variance were performed for each of the cognitive measures by hospital groupings to ensure that no differences in the cognitive processing indices of raters existed based on hospital groups. None were found.

Hypotheses

Hypothesis 1 - Trait versus Behavior Categories. Hypothesis One stated that raters will yield more accurate ratings to the extent that their category orientation, behavior or trait-based, corresponds to the orientation of the rating scale. Correlations between Behavior, Trait and the four accuracy measures are presented in Table 3. No significant correlations were found.

Since the hypothesis was stated as a more extreme either-or condition but tested by contiuous variables, three subgroups were formed. Raters' cognitive systems were classified as (1) behaviorally based if 75% of their responses were behavior constructs, (2) trait based if 75% of their responses were trait constructs, and (3) mixed if they did not fall into either of the first two groups. Four separate one-way analyses of variance were conducted for each of the four accuracy measures by the category orientation classification identified above. Again, no support was found for the effect of raters' category orientation on accuracy in ratings.

Hypothesis 2 - Degree of Match to Rating Scale. Hypothesis Two predicted that accuracy in ratings would be greater to the extent that raters were able to dimensionalize job behaviors in a manner consistent with the rating scale. In support of this hypothesis, correlations between rating scale match and both BARS accuracy measures were significant (for overall, $r = -.22$, $p = .007$ and, for correlational, $r = .27$, $p = .002$) and in the predicted direction.

Hypothesis 3 - Distinguishing Between Job Relevant and Non-Job Relevant Behaviors and Dimensions. Hypothesis Three stated that accuracy in ratings would be related to the degree to which a rater was able to distinguish between behaviors and dimensions which were relevant to job performance and those that were irrelevant to job performance. No significant relationships were found between Job Behavior Classification, Non-Job Behavior Classification, and BARS overall and correlational accuracy (see Table 3).

Hypothesis 4 - Cognitive Differentiation. Hypothesis Four posited that raters who were more accurate in their ratings would have more highly differentiated category systems for the job such that little overlap would exist between category dimensions while raters who provided less accurate ratings would be unable to differentiate clearly among dimensions. Correlational results showed that overall cognitive differentiation was only marginally related to correlational accuracy ($r = -.13$, $p = .08$, one-tailed). Interestingly, this effect was dependent on the type of behavior

dimensionalized by the rater. There were no significant correlations when differentiation was assessed for behaviors related to job performance; however, when nurses dimensionalized behaviors seen on the job but unrelated to job performance, a significant correlation resulted for correlational accuracy using the BARS scale ($r = -.18$, $p = .02$). This finding suggests that the better the rater was able to dimensionalize non-job related behaviors into dimensions with little degree of overlap, the more accurate were his or her ratings.

Hypothesis 5 - Cognitive Differentiation and Halo. Hypothesis Five proposed that raters with more highly differentiated category systems for the job would exhibit less halo in their ratings than those with less differentiated systems. The three measures of cognitive differentiation (Overall, Job and Non-Job) were each correlated with the two measures of halo--halo for the BARS scale and halo for the Trait scale. Results of these analyses are reported in Table 4. In support of the hypothesis, significant correlations were found between each of the cognitive differentiation measures and each of the halo measures.

It is also interesting to note that there were no significant correlations found between any of the above cognitive measures of dimensionalizing behaviors and/or rating scale match and either of the Trait accuracy measures. (The correlations ranged from .01 to .12, see Table 3). As expected, cognitive processes of raters assessed in the manners mentioned above appear unrelated to

Table 4

Correlations for Cognitive Differentiation Measures by Halo

	Cognitive Differentiation		
Halo	Overall	Job	Non-Job
Halo - Bars Scale	-.19[a]	-.18	-.19
Halo - Trait Scale	-.19	-.18	-.19

[a]Correlations above .18 are significant at $p \leq .05$ for two-tailed tests.

accuracy in ratings when using a trait-based rating scale. Our cognitive processing indices focused on behaviors rather than traits.

Hypothesis 6 - Experiences of the Rater. Hypothesis Six stated that experiences of the rater would correlate with the category system she or he used in evaluating the job performance of others and with rating accuracy. Because 97% of the nurses were female and because 91% were Registered Nurses, no analyses were performed based on sex or job title.

Correlations between experience variables and cognitive measures are reported in Table 5. The number of years worked as a nurse was negatively correlated with the degree to which nurses dimensionalized behaviors in a manner consistent with the rating scale (r = -.16, p = .04). The job position of the nurse was significantly related to several cognitive processing variables. The higher the job position of the nurse, the less likely she or he was to "miscategorize" behaviors by placing non-job related behaviors into job dimensions or job behaviors in non-job related dimensions (r = -.19, p = .03 and -.20, p = .02 respectively).

Additional analyses were performed to determine if any of the prior experiences of the rater were related to rating accuracy and are presented in Table 5. Only prior rating experience was significantly and positively correlated with any of the accuracy measures. Raters who had prior experience in rating nursing performance were more accurate in their ratings using the BARS

Table 5

Correlations of Experience Variables with Cognitive Variable and Accuracy

	Cognitive Variable									Accuracy Measures		
	Behavior	Trait	Dimensionalizing			Cognitive Differentiation						
Experience			Rating Scale Match	Nonjob Classification	Job Classification	Overall	Job	Nonjob	BARS Overall	BARS Correlational	Overall	Correlational
Years worked	.00	.00	-.16[a]	-.01	.07	-.01	-.03	.00	.06	.03	.11	-.01
Position	.08	-.03	-.04	-.19	-.20	-.14	-.12	-.15	-.02	-.04	.06	-.03
Schooling	-.09	.13	-.05	.07	-.04	.02	.07	-.04	-.09	.05	.03	-.02
Degree	-.08	.11	-.01	.04	-.03	.07	.10	.03	-.02	.03	.02	.07
Rating (yes or no)[b]	.00	-.01	-.03	-.08	-.09	-.09	-.08	-.10	.15	-.17	.10	-.04
Years rating	-.01	.01	-.02	.00	.05	.03	.02	.03	-.06	.10	-.01	.06

[a]Correlations above .15 (or -.15) depending upon the hypothesis being tested are significant at $p \leq .05$ for one-tailed tests and .18 for two-tailed tests.

[b]Point-biserial correlations.

scale than those without such prior rating experience (r = .15, p = .05 and -.17, p = .03). No significant results emerged for experiences of the rater in relation to rating accuracy using the Trait scale.

A one-way analysis of variance was also performed to determine if differences in rating accuracy were related to the unit in the hospital in which the nurse worked. Results indicated that differences in BARS correlational accuracy did exist by hospital unit ($F(7,116) = 3.57$, p = .002). Closer examination of the data revealed that nurses working in the surgery unit in the hospital were less accurate in their ratings for BARS correlational accuracy than those persons in any other hospital unit. No other differences by hospital unit were found for any of the other accuracy measures.

Overall, it appears that the variables of years worked as a nurse and job position were the important variables to consider for the cognitive processing of raters, while prior rating experience was important for rating accuracy using the BARS scale.

Discussion

Research and writing on performance appraisal theory and practices has shifted from a concern for the nature and form of rating scales and a description of appraisal practices to an attempt to understand the cognitive processes of the raters who complete such scales (Feldman, 1981; Ilgen & Feldman, 1983; Landy & Farr, 1980). The cognitive processing approach assumes that the

rater possesses some set of cognitive categories or "bins" in which information about others is stored and from which it is retrieved when the rater is asked to complete a performance rating of another. With the exception of Borman (1985), little or no research has attempted to focus directly on cognitive categories and their effects on ratings. The present research addressed this issue.

The results of this study supported, to some extent, the notion that cognitive categorization processes of raters are related to the accuracy of their performance ratings. Prior experience of raters was also investigated with respect to its relationship to cognitive processing variables and to rating accuracy. The results indicated that the amount of experience on the job and job position was related to the cognitive categories of the rater and that the amount of prior experience rating others was related to rating accuracy. The hypotheses addressed in this research cluster into three sets of issues which are addressed below.

Category Match to Scale

The first two hypotheses were predicated on the assumption that raters would be able to provide more accurate ratings of others the more their personal cognitive categories for storing information about others were consistent with the nature of the performance appraisal forms. Since extensive research on performance appraisal rating forms has lead to the conclusion that

information about ratee behaviors is preferred over information on traits, the first and more general hypothesis was that those who tended to use behavioral categories for storing information about others would be more accurate appraisers than those who stored information in more trait-like dimensions (Hypothesis 1). Likewise, when trait-focused scales were used, it was predicted that those who tended to have trait based views of others would be more accurate on trait scales than those with behavioral orientations. A more refined version of the matching hypothesis predicted that those whose cognitive categories matched the specific dimensions of the performance appraisal instrument used in the study would be more accurate in their ratings (Hypothesis 2).

The matching hypothesis was not supported at the general level, but did receive some support at the more specific level. When raters were given a list of job behaviors from the BARS scale, to be used later when rating performance, and were asked to sort the behaviors into the performance dimensions from the rating scale, those who were better able to sort the behaviors into the proper dimensions were also more accurate when using the BARS scale to rate performance. This finding is consistent with research on training for performance appraisal accuracy which has shown that accuracy improves when people are taught the performance dimensions and the behaviors comprising the dimensions prior to using the performance appraisal instruments (Bernardin & Pence, 1980; McIntyre, et al., 1984; Pulakos, 1984, in press). In the absense

of training, we found that the more accurate raters were those who already possessed a knowledge of the rating scale dimensions and the behaviors comprising those dimensions.

Several factors may have accounted for the failure of the more general trait or behavioral orientation of raters to differentiate good raters from poor ones. Given the more abstract level of the trait versus behavior orientation, as compared to the specific dimension match just discussed, we would expect the strength of the effect in the general condition to be weaker than the effect of the specific one. Since the specific hypothesis did hold up, but was not particularly strong, the strength of the specific relationship represented an upper bound for the more general one. The result was that the weaker general link was not observed.

Along similar lines, it is interesting to note that the general behavior-trait orientation, while not related to accuracy as hypothesized, was related to several of the cognitive processing indices. Specifically, behavior orientation was related to non-job cognitive differentiation (r = -.16, p = .04), while trait orientation correlated with non-job behavior classification (r = .19, p = .02) and cognitive differentiation (r = .18, p = .03). Marginally significant correlations of behavior orientation with non-job behavior classification (r = -.14, p = .06) and trait orientation with job behavior classification (r = .13, p = .07) and overall cognitive differentiation (r = .13, p = .07) were also consistent with these trends. These results seem to indicate that

the general category orientation of the rater may influence the more specific categories the rater develops for the job, which in turn, affect accuracy in rating job performance.

There is also the possibility that our trait versus behavior orientation measure did not reflect the categories people used to judge others. Certainly the ipsative nature of the measure made it impossible to address independently the effects of trait and behavior views. Although pretesting with the scale demonstrated reliable scores as trait or behaviorally focused, it was and is not a good criterion against which to assess whether those who score highly on trait (or behavior) orientation actually encode person perception information into trait (or behavior) categories. More work is needed on the trait-behavior hypothesis.

Category Precision

Several hypotheses were based on the asssumption that raters would differ in the extent to which they differentiated among categories used to judge others. With respect to appraisal accuracy, it was hypothesized that those with more differentiated category systems would by more accurate raters (Hypothesis 4) and would show lower levels of halo error (Hypothesis 5) than those who differentiated less among dimensions. Furthermore, it was believed that those with less precise category systems, reflected by the tendency to misclassify job behaviors into non-job dimensions and vice versa, would have less accurate performance ratings than those

with better ability to distinguish between relevant and irrelevant behaviors (Hypothesis 3).

The relationship of cognitive differentiation to rating accuracy appeared to depend on the type of behaviors dimensionalized. Specifically, only differentiation among non-job related behaviors was significantly related to rating accuracy. Feldman (1981) posits that when the rater is unable to clearly separate non-job related behaviors from job relevant information, the non-performance related behaviors will contribute to a general impression of the ratee that may bias ratings. Perhaps, in our case, those unable to differentiate non-job behaviors clearly did not perceive such behaviors in a multidimensional manner. The non-job behaviors may have been integrated into an overall general impression that was less accurate than one unaffected by the irrelevant behaviors.

Although there was some support for the fact that those higher in cognitive differentiation were more accurate when measured by the correlational accuracy index using the BARS scale, the strongest support for the differentiation hypothesis was found with respect to halo errors. In this case, those who differentiated more among dimensions for judging others had lower levels of halo in their responses on the performance appraisal instruments. To the extent that our cognitive differentiation measure reflects the level of cognitive complexity of the rater, these results support Bernardin et al.'s (1982) position that measures of cognitive

complexity that are relatively specific to the performance appraisal situation should be useful for examining rating accuracy and halo, and perhaps be more useful than the general complexity measures that are often employed in the literature.

Finally, no support was found for the hypothesis that those who misclassified job and non-job behaviors prior to rating others were less accurate raters than those who correctly classified behaviors. Several factors may have contributed to this. One of the most compelling reasons for the lack of support was the fact that the videotape used as the stimulus material almost exclusively focused on job related, rather than non-job related, behaviors. Thus, inaccuracies arising from cueing on non-job behaviors as contributors to performance were unlikely to occur due to the absense of such non-job behaviors in the stimulus materials. We would expect that in naturally occurring settings where non-job behaviors are much more prevalent, this issue may still be important.

Rater Experiences

Experience forms the basis for the development of cognitive categories (Rosch et al., 1976). Our data showed that experience was indeed related to cognitive category issues, but in some ways that were not initially anticipated. In particular, there was a negative relationship between the number of years of job experience and the ability to dimensionalize behaviors in a manner consistent with the rating scale. Although we expected the opposite, we

failed to consider the fact that over time, persons may have had experiences with performance appraisal instruments which were quite different from the ones used here. The greater the dissimilarity between our scale and those used by others, the more we would expect greater job experience to lead to greater divergence from the scale used in the study.

As expected, raters in higher job positions "miscategorized" fewer behaviors and also differentiated more between dimensions. The higher job positions may have influenced raters to develop different category systems, attend to different aspects of job performance and/or enable them to better distinguish between performance and non-performance related dimensions. A final interesting finding was that nurses' prior experience rating others was related to greater rating accuracy using the BARS scale. Perhaps, simple practice in making ratings enhanced accuracy.

Limitations of the Research

A great deal of care was taken in this research to develop ways to measure cognitive categories related to performance appraisal, create a videotaped set of stimuli that controlled the nature of the performance standard, and use subjects who were very familiar with the person's job being rated. From this we were able to find support for several of the hypotheses. In spite of this support, keep in mind that the relationships between cognitive category constructs and performance accuracy were not very strong. Yet, the relatively low level of relationship found is quite

consistent with the level observed in many other studies related to performance appraisal accuracy (see for example, Borman, 1977; Murphy et al. 1982). In all cases, a similar research paradigm was used in which participants were presented with a standard stimulus, primarily in the form of videotaped performance of a person or persons performing some task. The performance of the person on the tape was structured to represent the desired level of performance on preselected performance dimensions. Finally, expert judges who viewed the tape reached a relatively high level of agreement about the behaviors represented on the tape so that the standard possessed acceptable levels of reliability and validity.

In conducting the present study, a paradoxical dilemma surrounding this paradigm became apparent. On the one hand, research on performance appraisal accuracy requires the existence of some known standard to which ratings are compared. On the other hand, to create a standard with acceptably high agreement among expert judges requires that the performance behaviors and dimensions represented on the tape be very salient and obvious; only very clear behaviors survive the requirement for high rater agreement for the presence of the behavior in the stimulus materials. At the same time, those behaviors that are very obvious to the experts are also likely to be relatively obvious to the naive subjects. The result is that the requirements for a good standard (i.e., a "good stimulus tape") may greatly restrict the variance that can be observed in accuracy scores when the study is

conducted. (This same dilemma is also apparent when developing cognitive measures with acceptable reliability and validity.) To the extent that this is true, research that uses accuracy as a criterion is likely to find relatively low degrees of association between variables of interest. We are left with what appears to us to be a major limitation of the strength of past empirical research on performance appraisal accuracy and perhaps an unsolvable dilemma for future research on the topic using this commonly accepted experimental paradigm. The only encouraging conclusion gleaned from this is that some meaningful relationships have been found in this and other research on this topic. Given our belief that the method severely restricts the likelihood of observing relationships between selected variables and performance rating accuracy, we would expect that the effects observed in our restrictive setting would be much stronger in naturally occurring settings.

References

Adams-Webber, J. R. (1979). Personal Construct Theory. New York: John Wiley & Sons.

Bernardin, H. J., Cardy, R. L., & Carlyle, J. J. (1982). Cognitive complexity and appraisal effectiveness: Back to the drawing board? Journal of Applied Psychology, 67, 151-160.

Bernardin, H. J., & Pence, E. C. (1980). Effects of rater training: Creating new response sets and decreasing accuracy. Journal of Applied Psychology, 65, 60-66.

Bonarius, J. C. (1965). Research in the personal construct theory of George A. Kelly. In B. A. Maher (Ed.), Progress in experimental personality research Vol. 2. New York: Academic Press.

Borman, W. C. (1975). Effects of instructions to avoid halo error on reliability and validity of performance evaluation ratings. Journal of Applied Psychology, 60, 556-560.

Borman, W. C. (1977). Consistency of rating accuracy and rating errors in the judgment of human performance. Organizational Behavior and Human Performance, 20, 233-252.

Borman, W. C. (1978). Exploring upper limits of reliability and validity in job performance ratings. Journal of Applied Psychology, 63, 135-144.

Borman, W. C. (1979). Individual differences correlates of accuracy in evaluating others' performance effectiveness. Applied Psychological Measurement, 3, 103-115.

Borman, W. C. (1983). Implication of personality theory and research for the rating of work performance in organizations. In F. Landy, S. Zedeck, & J. Cleveland (Eds.), Performance Measurement and Theory. Hillsdale, NJ: Erlbaum. pp. 127-165.

Borman, W. C. (1985). Personal constructs, performance schema, and "folk theories" of subordinate effectiveness: Explorations in an army officer sample. Working paper.

Bruner, J. S., & Tagiuri, R. (1954). The perception of people. In G. Lindzey (Ed.), The handbook of social psychology Vol. 2. Reading, MA: Addison-Wesley.

Cooper, W. H. (1981). Ubiquitous halo. Psychological Bulletin, 90, 218-244.

Cronbach, L. J. (1955). Processes affecting scores on "understanding of others" and "assumed similarity." Psychological Bulletin, 52, 177-193.

Feldman, J. (1981). Beyond attribution theory: Cognitive processes in performance appraisal. Journal of Applied Psychology, 66, 127-148.

Hakel, M. (1969). Significance of implicit personality theories in personality research and theory. Proceedings of the 77th convention of the American Psychological Association. pp. 403-404.

Ilgen, D. R., & Feldman, J. M. (1983). Performance appraisal: A process approach. In B. M. Staw (Ed.), Research in organization behavior Vol. 2. Greenwich, CT: JAI Press.

Isaacson, G. I. (1966). A comparative study of meaningfulness of personal and common constructs. Unpublished dissertation, University of Missouri.

Kane, J. S., & Lawler, E. E., III. (1979). Performance appraisal effectiveness: Its assessments and determinants. In B. M. Staw (Ed.), Research in organizational behavior. Greenwich, CT: JAI Press.

Kelly, G. A. (1955). The psychology of personal constructs. New York: Norton.

Lahey, M. A., & Saal, F. E. (1981). Evidence incompatible with a cognitive compatibility theory of rating behavior. Journal of Applied Psychology, 6, 706-715.

Landy, F. J., & Farr, J. (1980). Performance rating. Psychological Bulletin, 87, 72-107.

McIntyre, R. M., Smith, D. E., & Hassett, C. E. (1984). Accuracy of performance ratings as affected by rater training and perceived purpose of rating. Journal of Applied Psychology, 69, 147-156.

Murphy, K. R., Martin, C., & Garcia, M. (1982). Do behavioral observation scales measure observation? Journal of Applied Psychology, 67, 562-567.

Nathan, B. R., & Alexander, R. A. (1985). The role of inferential accuracy in performance rating. Academy of Management Review, 10, 109-115.

Nathan, B. R., & Lord, R. G. (1983). Cognitive categorization and dimensional schemata: A process approach to the study of halo in performance ratings. Journal of Applied Psychology, 68, 102-114.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. Psychological Review, 84, 231-259.

Norman, W. T., & Goldberg, L. R. (1966). Raters, ratees, and randomness in personality structure. Journal of Personality and Social Psychology, 4, 681-691.

Ostroff, C. (1985). Cognitive categories of raters in performance appraisal and their relationship to rating accuracy. Unpublished masters thesis, Michigan State University, East Lansing.

Passini, F. T., & Norman, W. T. (1966). A universal conception of personality structure? Journal of Personality and Social Psychology, 4, 44-49.

Pulakos, E. D. (1984). A comparison of rater training programs: Error training versus accuracy training. Journal of Applied Psychology, 69, 581-588.

Pulakos, E. D. (in press). The development of training programs to increase accuracy with different rating tasks. Organizational Behavior and Human Decision Processes.

Rosch, E., Mervis, C. G., Gray, W. D., Johnson, D. M., & Boyers-Braem, P. (1976). Basic objects in natural categories. Cognitive Psychology, 8, 382-439.

Rosenberg, S. (1977). New approaches to the analysis of personal constructs in person perception. In Nebraska Symposium on Motivation 1976. Lincoln: University of Nebraska Press.

Sauser, W. I., & Pond, S. B. (1981). Effects of rater training and participation on cognitive complexity: An exploration of Schneier's cognitive reinterpretation. Personnel Psychology, 34, 563-577.

Schneider, D. J. (1973). Implicit personality theory: A review. Psychological Bulletin, 79, 294-309.

Schneier, E. C. (1977). Operational utility and psychometric characteristics of behavioral expectation scales: A cognitive reinterpretation. Journal of Applied Psychology, 62, 541-548.

Sechrest, L. B. (1968). Personal constructs and personal characteristics. Journal of Individual Psychology, 24, 162-166.

Smith, P. D., & Kendall, L. M. (1963). Retranslation of expectations: An approach to construction of unambiguous anchors for rating scales. Journal of Applied Psychology, 47, 149-155.

Shweder, R. A., & D'Andrade, R. G. (1980). The systematic distortion hypothesis. In R. A. Shweder (Ed.), Fallible judgment in behavioral research Vol. 4. San Francisco, CA: Jossey-Bass.

Author Notes

This paper was based on research sponsored, in part, by Contract No. N00014-83-K-0756 from the Office of Naval Research. Although support from the agency was greatly appreciated, the ideas expressed herein are those of the authors and not necessarily those of the supporting agency.

This article is based on the first author's master's thesis, directed by the second author. The authors would like to thank members of the thesis committee, Neal Schmitt and Steve Kozlowski, for their valuable contributions to the project.

Requests for reprints should be sent to Cheri Ostroff, Department of Psychology, Michigan State University, East Lansing, MI 48824-1117.

Footnotes

[1]These same nurses participated in two additional data collection sessions following the one described here, but only the instrument development and the first data collections from the nurses are described here.

[2]Dimensions eliminated primarily because of technical difficulties in filming the example behaviors for the dimensions in ways that produced high interrater agreement among experts who viewed the films.

LIST 1 MANDATORY*

Defense Technical Information Center (12)
ATTN: DTIC DDA-2
Selection & Preliminary Cataloging Section
Cameron Station
Alexandria, VA 22314

Library of Congress
Science and Technology Division
Washington, DC 20540

Office of Naval Research (3)
Code 4420E
800 N. Quincy Street
Arlington, VA 22217

Naval Research Laboratory (6)
Code 2627
Washington, DC 20375

Office of Naval Research
Director, Technology Programs
Code 200
800 N. Quincy Street
Arlington, VA 22217

LIST 2 ONR FIELD

Psychologist
Office of Naval Research
Detachment, Pasadena
1030 East Green Street
Pasadena, CA 91106

LIST 3 OPNAV

Deputy Chief of Naval Operations
(Manpower, Personnel & Training)
Head, Research, Development, and
Studies Branch (OP-115)
1812 Arlington Annex
Washington, DC 20350

Director
Civilian Personnel Division (OP-14)
Department of the Navy
1803 Arlington Annex
Washington, DC 20350

Deputy Chief of Naval Operations
(Manpower, Personnel, & Training)
Director, Human Resource Management
Plans & Policy Branch (OP-150)
Department of the Navy
Washington, DC 20350

LIST 4 NAVMAT & NPRDC

Program Administrator for Manpower,
Personnel, and Training
MAT-0722
800 N. Quincy Street
Arlington, VA 22217

Naval Material Command
Management Training Center
NAVMAT 09M32
Jefferson Plaza, Bldg #2, Rm 150
1421 Jefferson Davis Highway
Arlington, VA 20360

Naval Material Command
Director, Productivity Management
Office
MAT-00K
Crystal Plaza #5, Rm 632
Washington, DC 20360

Naval Personnel R&D Center (4)
Technical Director
Director, Manpower & Personnel
Laboratory, Code 06
Director, System Laboratory, Code 07
Director, Future Technology, Code 41
San Diego, CA 92152

*Number in parentheses is the number of copies to be sent.

Navy Personnel R&D Center
Washington Liaison Office
Ballston Tower #3, Rm 93
Arlington, VA 22217

LIST 5 BUMED

NONE

LIST 6
NAVAL ACADEMY AND NAVAL POSTGRADUATE SCHOOL

Naval Postgraduate School (3)
ATTN: Chairman, Dept. of
Administrative Science
Department of Administrative Sciences
Monterey, CA 93940

U.S. Naval Academy
ATTN: Chairman
Department of Leadership & Law
Stop 7-B
Annapolis, MD 21402

LIST 7 HRM

Officer in Charge
Human Resource Management Division
Naval Air Station
Mayport, FL 32228

Human Resource Management School
Naval Air Station Memphis (96)
Millington, TN 38054

Commanding Officer
Human Resource Management School
Naval Air Station Memphis
Millington, TN 38054

LIST 8 NAVY MISCELLANEOUS

Naval Military Personnel Command (2)
HRM Department (NMPC-6)
Washington, DC 20350

LIST 9 USMC

Headquarters, U.S. Marine Corps
ATTN: Scientific Adviser,
Code RD-1
Washington, DC 20380

LIST 10 OTHER FEDERAL GOVERNMENT

Dr. Brian Usilaner
GAO
Washington, DC 20548

Social and Developmental Psychology
Program
National Science Foundation
Washington, DC 20550

Office of Personnel Management
Office of Planning and Evaluation
Research Management Division
1900 E. Street, NW
Washington, DC 20415

LIST 11 ARMY

Technical Director (3)
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Head, Department of Behavior
Science and Leadership
U.S. Military Academy
New York 10996

LIST 12 AIR FORCE

Air University Library
LSE 76-443
Maxwell AFB, AL 36112

Head, Department of Behavioral
Science and Leadership
U.S. Air Force Academy
Colorado 80840

LIST 13 MISCELLANEOUS

Dr. Eduardo Salas
Human Factors Division
Code 712
Navy Training Systems Center
Department of the Navy
Orlando, FL 32813-7100

LIST 14 CURRENT CONTRACTORS

Dr. Janet L. Barnes-Farrell
Department of Psychology U-20
University of Connecticut
406 Cross Campus Road
Storrs, CT 06268

Jeanne M. Brett
Northwestern University
Graduate School of Management
2001 Sheridan Road
Evanston, IL 60201

Dr. Terry Connolly
Georgia Institute of Technology
School of Industrial & Systems
Engineering
Atlanta, GA 30332

Dr. Richard Daft
Texas A&M University
Department of Management
College Station, TX 77843

Dr. Randy Dunham
University of Wisconsin
Graduate School of Business
Madison, WI 53706

Dr. Lawrence R. James
School of Psychology
Georgia Institute of Technology
Atlanta, GA 30332

Dr. J. Richard Hackman
School of Organization & Management
Box 1A
Yale University
New Haven, CT 06520

Dr. Frank J. Landy
Department of Psychology
Pennsylvania State University
450 Moore Bldg.
University Park, PA 16802

Dr. Bibb Latane
University of North Carolina
at Chapel Hill
Manning Hall 026A
Chapel Hill, NC 27514

Dr. Edward E. Lawler III
Graduate School of Business
University of Southern California
Los Angeles, CA 90007

Dr. William H. Mobley
College of Business Administration
Texas A&M University
College Station, TX 77843

Dr. Thomas M. Ostrom
The Ohio State University
Department of Psychology
404C West 17th Avenue
Columbus, OH 43210

Dr. Robert Rice
Department of Psychology
SUNY - Buffalo
Buffalo, NY 14226

Dr. Benjamin Schneider
Department of Psychology
University of Maryland
College Park, MD 20742

Dr. H. Wallace Sinaiko
Program Director, Manpower Research
and Advisory Services
Smithsonian Institution
801 N. Pitt Street, Suite 120
Alexandria, VA 22314

Dr. Richard M. Steers
Graduate School of Management
University of Oregon
Eugene, OR 97403

Dr. Harry C. Triandis
Department of Psychology
University of Illinois
Champaign, IL 61820

Dr. Anne S. Tsui
Duke University
The Fuqua School of Business
Durham, NC 27706

Andrew H. Van de Ven
University of Minnesota
Office of Research Administration
1919 University Avenue
St. Paul, MN 55104

END

FILMED

1-86

DTIC